The Information Matrix for Image Factor Analysis

M. W. Browne

Department of Statistics
University of South Africa
P.O. Box 392
0001 Pretoria, South Africa

Submitted by George P. H. Styan

ABSTRACT

The information matrix and gradient of the likelihood function are derived for the image factor analysis model using methods of matrix differentiation.

1. INTRODUCTION

The traditional factor analysis (TFA) model is given by

$$\Sigma = \Lambda\Phi\Lambda' + D_\psi, \tag{1}$$

where a value assumed by the $p \times p$ symmetric matrix valued function $\Sigma = \Sigma(\Lambda, \Phi, D_\psi)$ represents a covariance matrix, Λ represents a $p \times k$ factor matrix, Φ represents a $k \times k$ symmetric factor covariance matrix, and D_ψ is a diagonal matrix representing unique variances. Certain elements of Λ, Φ, and D_ψ are fixed, so that the remaining parameters will be identified. Suppose that the sample covariance matrix S has a Wishart distribution with n degrees of freedom and expected value $\Sigma_0 = \Sigma(\Lambda_0, \Phi_0, D_{\psi_0})$. Maximum likelihood (ML) estimates $\hat{\Lambda}$, $\hat{\Phi}$, and $\hat{D}_\psi$, may be obtained by minimizing the discrepancy function [8, 7]

$$F = \log|\Sigma| - \log|S| + \operatorname{tr}[S\Sigma^{-1}] - p \tag{2}$$

with respect to the free elements of Λ, Φ, and D_ψ using a pseudo-Newton algorithm. Lawley [8] and Jöreskog [7] give the gradient vector required and an approximation, $H = H(\Lambda, \Phi, D_\psi)$, to the Hessian with the property that

52 Vanderbilt Ave., New York, NY 10017 0024-3795/85/$3.30

$H(\Lambda_0, \Phi_0, D_{\psi_0})$ is equal to the expected value of the Hessian when Σ_0 satisfies the model. The information matrix is given by $\frac{1}{2}nH$, so that the asymptotic covariance matrix of the maximum likelihood estimators is given by $2n^{-1}(H(\Lambda_0, \Phi_0, D_{\psi_0}))^{-1}$. Jöreskog [7] employed the Fletcher-Powell [3] algorithm to minimize F. He made use of the expected Hessian, H, to provide an initial approximation for the Hessian and thereby improve the speed of convergence of the algorithm. An alternative algorithm, which is recommended by Lee and Jennrich [9], is the Fisher scoring algorithm in which the Hessian of the usual Newton-Raphson algorithm is replaced by the expected Hessian H.

Jöreskog [5, 6] suggested a modification of the TFA model (1) motivated by Guttman's [4] image theory, which he called the image factor analysis (IFA) model. This is given by

$$\Sigma = \Lambda\Phi\Lambda' + \theta(\mathrm{Diag}[\Sigma^{-1}])^{-1}, \tag{3}$$

where $\mathrm{Diag}[X]$ stands for a diagonal matrix formed from the diagonal elements of a square matrix X. Jöreskog [6] provided a rationale for the IFA model and pointed out advantages of the IFA model over the TFA model.

The IFA model (3) differs from the TFA model (1) in that the diagonal elements of Σ in (3) are implicit functions of the parameters Λ, Φ, and θ, rather than explicit functions. Jöreskog [6] obtained the gradient of the ML discrepancy function F with respect to the parameters of the IFA model by differentiating a Lagrangian function. This approach, however, did not lend itself to obtaining the expected Hessian H, so that the identity matrix had to be employed to initiate the Fletcher-Powell procedure, and the asymptotic covariance matrix of ML estimators was not available.

The purpose of the present paper is to obtain the expected Hessian, and consequently the information matrix, for the IFA model. The approach employed here differs from that of Jöreskog. Some general results, due to Bargmann [1], which express the elements of the gradient and information matrix in terms of partial derivatives of Σ, are applied to provide an alternative derivation of the gradient, first given by Jöreskog, and the expected Hessian.

2. PRELIMINARY RESULTS

Let $\Sigma(\gamma)$ represent a twice continuously differentiable symmetric matrix valued function of a $q \times 1$ parameter vector γ. The symmetric matrix,

$\partial\Sigma/\partial\gamma_i$, of partial derivatives of the elements of Σ with respect to the ith element γ_i of $\boldsymbol{\gamma}$ will be represented alternatively by $\dot{\Sigma}_i$ or by $\dot{\Sigma}\{\gamma_i\}$. In general a quantity enclosed in braces, { }, will be regarded as an indexing quantity rather than as an argument of a function. Thus $\dot{\Sigma}\{\gamma_i\}$ is that symmetric matrix valued function of $\boldsymbol{\gamma}$ which results from the differentiation of Σ with respect to γ_i.

The following result, given by Bargmann [1], simplifies the derivation of maximum likelihood estimators and the information matrix for arbitrary structural models for covariance matrices, and will be useful here.

LEMMA 1. *If F is the ML discrepancy function given in* (2), *the gradient of F has the typical element*

$$\frac{\partial F}{\partial \gamma_i} = -\operatorname{tr}[G\dot{\Sigma}_i] \tag{4}$$

where

$$G = \Sigma^{-1}(S - \Sigma)\Sigma^{-1}, \tag{5}$$

and the expected value of the Hessian of F has elements given by

$$h_{ij} = h\{\gamma_i, \gamma_j\} = \mathscr{E}\left(\frac{\partial^2 F}{\partial\gamma_i\,\partial\gamma_j}\right) = \operatorname{tr}\left[\Sigma^{-1}\dot{\Sigma}_i\Sigma^{-1}\dot{\Sigma}_j\right] \tag{6}$$

at $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$.

The parameter vector $\boldsymbol{\gamma}$ here consists of θ and the free elements of Λ and Φ.

Let $\boldsymbol{\alpha}$ be the vector of diagonal elements of Σ^{-1}, and let

$$D_\alpha = \operatorname{Diag}[\Sigma^{-1}], \tag{7}$$

so that the IFA model (3) may be written as

$$\Sigma = \Lambda\Phi\Lambda' + \theta D_\alpha^{-1}. \tag{8}$$

Clearly the nondiagonal elements of Σ are twice continuously differentiable functions of $\boldsymbol{\gamma}$. Let $[X]_{ij}$ represent the element in the ith row and jth column of a matrix X, and let $(X)^{(k)}$ be a matrix obtained from X by raising each element to the kth power, i.e. $[(X)^{(k)}]_{ij} = ([X]_{ij})^k$. Use of the implicit function theorem will show (cf. [6, Theorem 6]) that, locally, the diagonal elements of Σ are also twice continuously differentiable implicit functions of $\boldsymbol{\gamma}$ provided that the symmetric matrix

$$Q = \theta^{-1} D_\alpha^2 - (\Sigma^{-1})^{(2)} \tag{9}$$

is nonsingular.

Lemma 2. *If Q is nonsingular, then*

$$\dot{\Sigma}_i = \dot{\Sigma}\{\gamma_i\} = B_i + D_{u_i} \tag{10}$$

where

$$B_i = B\{\gamma_i\} = \frac{\partial(\Lambda\Phi\Lambda')}{\partial\gamma_i} + \frac{\partial\theta}{\partial\gamma_i} D_\alpha^{-1} \tag{11}$$

is the derivative of Σ with respect to γ_i disregarding the fact that D_α is a function of $\boldsymbol{\gamma}$, and the vector of diagonal elements of D_{u_i} is given by

$$\mathbf{u}_i = Q^{-1}\mathbf{v}_i, \tag{12}$$

where

$$\mathbf{v}_i = \mathbf{v}\{\gamma_i\} = \operatorname{diag}\left[\Sigma^{-1} B_i \Sigma^{-1}\right], \tag{13}$$

and $\operatorname{diag}[X]$ represents a vector formed from the diagonal elements of a square matrix X.

Proof. Differentiation of both sides of (8) with respect to γ_i yields

$$\dot{\Sigma}_i = B_i + \theta D_\alpha^{-2} \operatorname{Diag}\left[\Sigma^{-1}\dot{\Sigma}_i\Sigma^{-1}\right].$$

Let

$$D_{u_i} = \operatorname{Diag}[\dot{\Sigma}_i] - \operatorname{Diag}[B_i], \tag{14}$$

so that

$$D_{u_i} = \theta D_\alpha^{-2} \operatorname{Diag}\left[\Sigma^{-1}(B_i + D_{u_i})\Sigma^{-1}\right]. \tag{15}$$

Taking the vectors of diagonal elements of both sides of (15) yields

$$\left(\theta^{-1}D_\alpha^2 - (\Sigma^{-1})^{(2)}\right)\mathbf{u}_i = \mathrm{diag}\left[\Sigma^{-1}B_i\Sigma^{-1}\right]. \tag{16}$$

Equation (12) now follows from (16), and (10) from (14). ■

3. THE GRADIENT AND EXPECTED HESSIAN FOR THE IFA MODEL

The expressions for $\partial F/\partial\gamma_i$ and $h\{\gamma_i, \gamma_j\}$ when $\Sigma(\boldsymbol{\gamma})$ is any covariance structure, given in Lemma 1, may be specialized to the IFA structure with the use of Lemma 2.

LEMMA 3. *Suppose that* γ_i *and* γ_j *are any parameters in the IFA structure given in* (8) *and* (7). *Then*

$$\frac{\partial F}{\partial\gamma_i} = -\mathrm{tr}\left[\Sigma^{-1}(S - \Sigma + D_z)\Sigma^{-1}B_i\right], \tag{17}$$

where

$$\mathbf{z} = Q^{-1}\mathrm{diag}[G], \tag{18}$$

and

$$h\{\gamma_i, \gamma_j\} = \mathrm{tr}\left[\Sigma^{-1}B_i\Sigma^{-1}B_j\right] + \mathbf{v}_i'W\mathbf{v}_j \tag{19}$$

where

$$W = Q^{-1}\left[2\theta^{-1}D_\alpha^2 - (\Sigma^{-1})^{(2)}\right]Q^{-1}. \tag{20}$$

Proof. Substitution of (10) into (4) yields

$$\frac{\partial F}{\partial\gamma_i} = -\mathrm{tr}[GB_i] - \mathrm{tr}\left[GD_{u_i}\right]. \tag{21}$$

But, if $\mathbf{g} = \mathrm{diag}[G]$, use of (12), (13), and (18) shows that

$$\begin{aligned}\mathrm{tr}\left[GD_{u_i}\right] &= \mathbf{g}'\mathbf{u}_i = \mathbf{g}'Q^{-1}\mathbf{v}_i = \mathbf{z}'\mathrm{diag}\left[\Sigma^{-1}B_i\Sigma^{-1}\right] \\ &= \mathrm{tr}\left[\Sigma^{-1}D_z\Sigma^{-1}B_i\right]. \end{aligned} \tag{22}$$

Equation (17) now follows from (21) and (22).

Substitution of (10) into (6) and use of (12) and (13) yields

$$h\{\gamma_i,\gamma_j\} = \mathrm{tr}\left[\Sigma^{-1}B_i\Sigma^{-1}B_j\right] + \mathrm{tr}\left[\Sigma^{-1}B_i\Sigma^{-1}D_{u_j}\right] + \mathrm{tr}\left[\Sigma^{-1}B_j\Sigma^{-1}D_{u_i}\right]$$

$$+\,\mathrm{tr}\left[\Sigma^{-1}D_{u_i}\Sigma^{-1}D_{u_j}\right]$$

$$= \mathrm{tr}\left[\Sigma^{-1}B_i\Sigma^{-1}B_j\right] + \mathbf{v}_i'\mathbf{u}_j + \mathbf{v}_j'\mathbf{u}_i$$

$$+\,\mathbf{u}_i'(\Sigma^{-1})^{(2)}\mathbf{u}_j$$

$$= \mathrm{tr}\left[\Sigma^{-1}B_i\Sigma^{-1}B_j\right]$$

$$+\,\mathbf{v}_i'\left(2Q^{-1} + Q^{-1}(\Sigma^{-1})^{(2)}Q^{-1}\right)\mathbf{v}_j. \tag{23}$$

Equation (19) now follows from (23) after use of (20) and (9). ■

Specific elements of the gradient and expected Hessian may now be obtained from the general results given in Lemma 3.

THEOREM 1. *The partial derivatives of the ML discrepancy function F in* (2) *with respect to* θ *and the elements of* Λ *and* Φ *in the IFA model are given by*

$$\frac{\partial F}{\partial[\Lambda]_{ir}} = -2\left[\Sigma^{-1}(S-\Sigma+D_z)\Sigma^{-1}\Lambda\Phi\right]_{ir}, \tag{24}$$

$$\frac{\partial F}{\partial[\Phi]_{rs}} = -(2-[I]_{rs})\left[\Lambda'\Sigma^{-1}(S-\Sigma+D_z)\Sigma^{-1}\Lambda\right]_{rs}, \tag{25}$$

$$\frac{\partial F}{\partial\theta} = -\mathrm{tr}\left[\Sigma^{-1}(S-\Sigma+D_z)\Sigma^{-1}D_\alpha^{-1}\right]. \tag{26}$$

If $[X]_{\cdot i}$ *stands for a column vector formed from the ith column of a matrix X and* $[X]_{\cdot i} * [Y]_{\cdot r}$ *represents the Hadamard product of the vectors* $[X]_{\cdot i}$ *and* $[Y]_{\cdot r}$ *with its tth element given by* $[[X]_{\cdot i} * [Y]_{\cdot r}]_t = [X]_{ti}[Y]_{tr}$, *the elements of the expected Hessian H are*

$$h\{[\Lambda]_{ir},[\Lambda]_{js}\} = 2\left([\Sigma^{-1}]_{ij}[\Phi\Lambda'\Sigma^{-1}\Lambda\Phi]_{rs} + [\Sigma^{-1}\Lambda\Phi]_{is}[\Sigma^{-1}\Lambda\Phi]_{jr}\right)$$

$$+\,4\left([\Sigma^{-1}]_{\cdot i} * [\Sigma^{-1}\Lambda\Phi]_{\cdot r}\right)'W\left([\Sigma^{-1}]_{\cdot j} * [\Sigma^{-1}\Lambda\Phi]_{\cdot s}\right), \tag{27}$$

$$h\{[\Lambda]_{ir},[\Phi]_{tu}\} = (2-[I]_{tu})\big([\Sigma^{-1}\Lambda]_{it}[\Lambda'\Sigma^{-1}\Lambda\Phi]_{ur} + [\Sigma^{-1}\Lambda]_{iu}[\Lambda'\Sigma^{-1}\Lambda\Phi]_{tr}\big) + 2(2-[I]_{tu})\big([\Sigma^{-1}]_{\cdot i} * [\Sigma^{-1}\Lambda\Phi]_{\cdot r}\big)' \times W\big([\Sigma^{-1}\Lambda]_{\cdot t} * [\Sigma^{-1}\Lambda]_{\cdot u}\big), \tag{28}$$

$$h\{[\Lambda]_{ir},\theta\} = 2\big[\Sigma^{-1}D_\alpha^{-1}\Sigma^{-1}\Lambda\Phi\big]_{ir} + 2\big([\Sigma^{-1}]_{\cdot i} * [\Sigma^{-1}\Lambda\Phi]_{\cdot r}\big)'W(\Sigma^{-1})^{(2)}(\alpha)^{(-1)}, \tag{29}$$

$$h\{[\Phi]_{rs},[\Phi]_{tu}\} = \tfrac{1}{2}(2-[I]_{rs})(2-[I]_{tu})\big([\Lambda'\Sigma^{-1}\Lambda]_{rt}[\Lambda'\Sigma^{-1}\Lambda]_{su} + [\Lambda'\Sigma^{-1}\Lambda]_{ru}[\Lambda'\Sigma^{-1}\Lambda]_{st}\big) + (2-[I]_{rs})(2-[I]_{tu}) \times \big([\Sigma^{-1}\Lambda]_{\cdot r} * [\Sigma^{-1}\Lambda]_{\cdot s}\big)'W\big([\Sigma^{-1}\Lambda]_{\cdot t} * [\Sigma^{-1}\Lambda]_{\cdot u}\big), \tag{30}$$

$$h\{[\Phi]_{rs},\theta\} = (2-[I]_{rs})\big[\Lambda'\Sigma^{-1}D_\alpha^{-1}\Sigma^{-1}\Lambda\big]_{rs} + (2-[I]_{rs})\big([\Sigma^{-1}\Lambda]_{\cdot r} * [\Sigma^{-1}\Lambda]_{\cdot s}\big)'W(\Sigma^{-1})^{(2)}(\alpha)^{(-1)}, \tag{31}$$

$$h\{\theta,\theta\} = (\alpha')^{(-1)}\big((\Sigma^{-1})^{(2)} + (\Sigma^{-1})^{(2)}W(\Sigma^{-1})^{(2)}\big)(\alpha)^{(-1)}. \tag{32}$$

Proof. Let J_{ij} represent a matrix with null elements except for the element in the ith row and jth column, which is unity. It follows from (11) that

$$B\{[\Lambda]_{ir}\} = J_{ir}\Phi\Lambda' + \Lambda\Phi J_{ri}, \tag{33}$$

$$B\{[\Phi]_{rs}\} = \tfrac{1}{2}(2-[I]_{rs})\Lambda(J_{rs} + J_{sr})\Lambda', \tag{34}$$

$$B\{\theta\} = D_\alpha^{-1}. \tag{35}$$

Substitution of (33), (34), (35) into (17) and use of trace manipulation rules [1, Section 7] yields (24), (25), (26). Further substitution of (33), (34), (35) into (13) and (19) yields (27), (28), (29), (30), (31), (32). ■

Jöreskog [6, Equations (23), (24)] has given expressions for the gradient which are equivalent to (24) and (26) with $\Phi = I$. The matrices $D_{\tilde{z}}$ in (24)–(26) and W in (27)–(32) appear in correction terms for (7). If $D_{\tilde{z}}$ and W are replaced by null matrices in Theorem 1, expressions are obtained which are appropriate for the covariance structure defined by (8) with $D_{\tilde{\alpha}}$ regarded as a fixed matrix. Similar methods to those employed by Jöreskog [7] may be used to carry out the computations involved in Theorem 1 effectively.

The gradient and expected Hessian with typical elements given in Theorem 1 may alternatively be expressed in matrix notation. Let Vec$[X]$ represent a $p^2 \times 1$ vector formed by stacking the columns of a $p \times p$ matrix X one on top of the other (e.g. [10, Section 4.2]), and let Δ be a $p^2 \times q$ matrix with columns formed from the elements of the B matrices defined in (33)–(35), i.e. $[\Delta]_{\cdot j} = \mathrm{Vec}[B\{\gamma_j\}]$. Then the $q \times 1$ gradient of F is given by

$$\frac{\partial F}{\partial \gamma} = -\Delta'(\Sigma^{-1} \otimes \Sigma^{-1})\mathrm{Vec}[S - \Sigma + D_{\tilde{z}}], \tag{36}$$

where $\Sigma^{-1} \otimes \Sigma^{-1}$ is the Kronecker product of Σ^{-1} with Σ^{-1}.

Let P be the $p^2 \times p$ transition matrix [10, p. 160] which takes the $p^2 \times 1$ vector Vec$[X]$ into the $p \times 1$ vector diag$[X]$, i.e. $P'\,\mathrm{Vec}[X] = \mathrm{diag}[X]$. A typical element of P is given by [2, Section 2] $[P]_{ij,g} = [I]_{ig}[I]_{jg}$. Then

$$H = \Delta'(\Sigma^{-1} \otimes \Sigma^{-1})\Delta + \Delta'(\Sigma^{-1} \otimes \Sigma^{-1})PWP'(\Sigma^{-1} \otimes \Sigma^{-1})\Delta. \tag{37}$$

Replacement of $D_{\tilde{z}}$ and W in (36) and (37) by null matrices results in familiar expressions for the gradient and expected Hessian of the ML discrepancy function under the assumption of a general covariance structure (cf. [2]).

Helpful discussions with Alexander Shapiro are acknowledged with thanks.

REFERENCES

1 R. E. Bargmann, Matrices and determinants, in *Chemical Rubber Company Handbook of Tables for Mathematics* (S. M. Selby, Ed.), Chemical Rubber Company, 1967.
2 M. W. Browne, Generalized least squares estimators in the analysis of covariance structures, *South African Statist. J.* 8:1–24 (1974).
3 R. Fletcher and M. J. D. Powell, A rapidly convergent descent method for minimization, *Comput. J.* 2:163–168 (1963).
4 L. Guttman, Image theory for the structure of quantitative variates, *Psychometrika* 18:277–296 (1953).

5 K. G. Jöreskog, *Statistical Estimation in Factor Analysis*, Almqvist and Wiksell, Stockholm, 1963.
6 K. G. Jöreskog, Efficient estimation in image factor analysis, *Psychometrika* 34:51–75 (1969).
7 K. G. Jöreskog, A general approach to confirmatory maximum likelihood factor analysis, *Psychometrika* 34:183–202 (1969).
8 D. N. Lawley, Some new results in maximum likelihood factor analysis, *Proc. Roy. Soc. Edinburgh, Sect. A* 67:256–264 (1969).
9 S. Y. Lee and R. I. Jennrich, A study of algorithms for covariance structure analysis with specific comparisons using factor analysis, *Psychometrika* 44:99–113 (1979).
10 D. G. Nel, On matrix differentiation in statistics, *South African Statist. J.* 14:137–193 (1980).

Received 18 April 1984; revised 9 November 1984